

# Detecting Apathy in Older Adults with Cognitive Disorders Using Automatic Speech Analysis

Alexandra König<sup>a,b</sup>, Nicklas Linz<sup>c</sup>, Radia Zeghari<sup>a</sup>, Xenia Klinge<sup>c</sup>, Johannes Tröger<sup>c</sup>, Jan Alexandersson<sup>c</sup> and Philippe Robert<sup>a</sup>

<sup>a</sup>*CoBTeK (Cognition-Behaviour-Technology) Lab, Memory Center CHU, Université Côte d'Azur, Nice, France*

<sup>b</sup>*INRIA Stars Team, Sophia Antipolis, Valbonne, France*

<sup>c</sup>*German Research Center for Artificial Intelligence (DFKI), Saarbrücken, Germany*

Accepted 16 April 2019

## Abstract.

**Background:** Apathy is present in several psychiatric and neurological conditions and has been found to have a severe negative effect on disease progression. In older people, it can be a predictor of increased dementia risk. Current assessment methods lack objectivity and sensitivity, thus new diagnostic tools and broad-scale screening technologies are needed.

**Objective:** This study is the first of its kind aiming to investigate whether automatic speech analysis could be used for characterization and detection of apathy.

**Methods:** A group of apathetic and non-aphetic patients ( $n=60$ ) with mild to moderate neurocognitive disorder were recorded while performing two short narrative speech tasks. Paralinguistic markers relating to prosodic, formant, source, and temporal qualities of speech were automatically extracted, examined between the groups and compared to baseline assessments. Machine learning experiments were carried out to validate the diagnostic power of extracted markers.

**Results:** Correlations between apathy sub-scales and features revealed a relation between temporal aspects of speech and the subdomains of reduction in interest and initiative, as well as between prosody features and the affective domain. Group differences were found to vary for males and females, depending on the task. Differences in temporal aspects of speech were found to be the most consistent difference between apathetic and non-aphetic patients. Machine learning models trained on speech features achieved top performances of AUC = 0.88 for males and AUC = 0.77 for females.

**Conclusions:** These findings reinforce the usability of speech as a reliable biomarker in the detection and assessment of apathy.

**Keywords:** Apathy, assessment, machine learning, neuropsychiatric symptoms, speech analysis, voice analysis

## INTRODUCTION

Apathy can be described generally as a syndrome comprising a reduction in goal-directed behaviors,

reduction of interests, and emotional blunting [1]. Study findings suggest that a disruption of mechanisms underlying the way in which reward is processed to motivate behavior could be the potential cause [2]. Consequently, it can be seen primarily as a motivational disorder present in several psychiatric and neurological conditions such as traumatic brain injury [3], major depression [4], or schizophrenia [5], as well as in neurodegenerative diseases including Alzheimer's disease (AD) [6] or Parkinson's disease

\*Correspondence to: Dr. Alexandra König, CoBTeK (Cognition, Behaviour, Technology) Lab Université Côte d'Azur, France; Centre Mémoire de Ressources et de Recherche, CHU de Nice, Institut Claude Pompidou, 10 rue Molière 06100, Nice, France. Tel.: +33 (0) 492 034 760; Fax: +33 (0) 4 93 52 92 57; E-mail: alexandra.konig@inria.fr.

[7]. Although there seems to be a lack of consensus in the definition across different pathologies, with different terms employed interchangeably according to patient groups, Cathomas et al. [8] proposed that for research purposes it may be helpful to regard it as one concept to a large extent, applicable across traditional nosological categories, to be considered a “trans-diagnostic clinical phenotype”.

The presence of apathy visibly and significantly affects the patient’s and caregiver’s quality of life [9]. In neurodegenerative disorders, apathy is associated with faster cognitive and functional decline [10] representing a risk factor for the conversion from early stages to AD. Thus, identifying apathy timely in disease progression is considered a clinical and research priority.

Current assessment methods for apathy rely mostly on scales or interview-based self-reports such as the Apathy Inventory [11] or the Neuropsychiatric Inventory [12], which might not always capture the actual state of a person’s level of motivation and activity since it is limited to the moment the patient is being evaluated. Furthermore, their application for early detection is rather limited because of their dependency on human observers as well as frequently impaired capacity for self-observation [13]. Thus, broad apathy measures may not sufficiently detect subtle variations in the presentation of apathy pointing to a need for additional more sensitive and objective assessment tools. Recently, a task force revised the apathy diagnostic criteria for better operationalization in clinical and research practice, stipulating the presence of quantitative reduction of goal-directed activity either in the behavioral, cognitive, emotional, or social dimension in comparison to the patient’s previous level of functioning [14]. With this, it was suggested that, information and communications technologies (ICT) might supplement these classical tools with additional objective measures, potentially providing more continuous endpoints in clinical trials. Several attempts have been made to investigate the use of ICT for apathy assessment over the past years. König et al. [15] performed a review of ICT for the assessment of apathy and concluded that no one had previously used ICT specifically in this context, but that techniques seemed promising. Since apathy seems to affect emotion-based decision-making, attempts to measure it through video games were made such as the Philadelphia Apathy Computerized Task (PACT) [16], detecting impairments in goal-directed behavior including initiation, planning, and motivation. Reward and effort mechanisms have

been explored along with physical effort discounting through paradigms such as the one developed by Pessiglione et al. [17]. Studies in schizophrenia have shown that actigraphy and the measurement of motor activity provide a promising readout for quantifying apathy [18]. Actigraphy has been used as well to measure physical changes in dementia patients with apathy [19]. Apathy has also been explored using eye-tracking in AD patients with the result that apathetic patients tend to fixate less on social images than non-apathetic patients [20]. Despite these efforts to find alternative objective measurements of apathy, an easy to implement, cheap and fast method, which could help with early, non-intrusive and potentially remote screening, is still urgently needed.

Recent advances in computational linguistics and language processing have led to the use of automatic speech analysis in the assessment of various clinical manifestations [21]. Semantic and acoustic features automatically extracted from speech tasks seemed highly sensitive to cognitive decline and potential conversion risk to dementia [22].

Significant associations were found between negative symptoms in schizophrenic patients and variability in pitch and speech proportion, even in different languages [23]. Strong correlations were obtained between negative symptom severity and phonetically measured reductions in tongue movements during speech [24].

Apathy belongs to the negative symptomatology of schizophrenia [25]. Recent factor analysis-based studies show the distinction of two underlying subdomains of negative symptoms, namely amotivation and diminished expression [26, 27]. Both domains may be characterized by aprosody (flattened speech intonation) and poor speech production (similar to the ‘emotional blunting’ domain in apathy) which seem to be easily detectable by this technology thus, making it a promising tool for measuring and tracking severity of symptoms, even across different types of population [28].

In depression, it is notable by ear that patients show a reduced prosody spectrum and sound rather monotonous which could serve as an indicator, if objective measurements can quantify these observations. Until now, several groups investigated the use of automatic analysis of speech as an additional assessment tool with an extensive review published by Cummins et al. [29] outlining the interest of using speech as a key objective marker for disease progression. Prosodic, articulatory, and acoustic features of speech seem affected by depression severity and thus

can easily be identified and used for continuously monitoring patients. With a considerable overlap of symptoms between depression and apathy, namely the lack of interest and goal-oriented behavior, we anticipate similar results when applying speech technology methods to apathy with a slightly different pattern in regards to emotionally triggered speech. To the best of our knowledge, at present, no other study aims to detect apathy by the means of automatic speech analysis. Therefore, the current study intends to investigate the feasibility of automatic analysis utilizing paralinguistic speech features extracted during a short free speech task as a potential candidate for clinical apathy assessment (characterization) and broad screening (detection) in elderly patients with cognitive impairment.

## METHODS

### *Participants*

60 patients aged 65 or older with neurocognitive disorder according to the Diagnostic Statistical Manual 5 (DSM-5) [30] were included in this study. For this, the presence of cognitive decline in memory and/or executive function with or without interference with independence was required based on previously performed evaluations. Participants underwent a clinical assessment including, among others, the Mini-Mental State Examination (MMSE) [31], the Apathy Inventory (AI) [11], and the Neuropsychiatric Inventory (NPI) [12]. Apathy was diagnosed based on the AI total score ( $\geq 4$ ). According to this assessment, participants were categorized into either non-apathy ( $N=30$ ) or apathy ( $N=30$ ) groups and matched for age and MMSE per gender group. In this study, we only look at patients with neurocognitive impairments, to prevent confounding of group differences by cognitive state. Thus, patients were selected from a larger existing cohort to build two comparable groups.

Speech features vary naturally between males and females. These differences have been leveraged in gender classification through speech analysis based on pitch and formant frequencies [32], Harmonic-to-Noise ratio [33], and linear predictive components and mel frequency cepstral coefficients (MFCC) [34]. Previous work found differences in speech depending on gender in the effects of apathy [35], as well as depression and the effectiveness of classifiers for its detection [36]. This is why this study considers males and females separately. All participants

were recruited through the Memory Clinic located at the Institute Claude Pompidou in the Nice University Hospital. Participants were all native speakers of French and excluded if they had any major auditory or language problems, history of head trauma, loss of consciousness, psychotic or aberrant motor behavior, or history of drug abuse. Written informed consent was obtained from all subjects prior to the experiments. The study was approved by Nice Ethics Committee (ELEMENT ID RCB 2017-A01896-45, MoTap ID RCB 2017-A01366-47) and was conducted according to the Declaration of Helsinki.

### *Speech protocol*

Free and natural speech tasks require low cognitive effort and are capable of eliciting emotional reactions (or a lack thereof) by asking to describe events that triggered recent affective arousal [37]. To this end, people were asked to perform two tasks: 1) talk about a positive event in their life and 2) to talk about a negative event in their life. Instructions (“*Can you tell me in one minute about a positive/negative event?/Pouvez-vous me raconter en une minute d’un événement positif/négatif?*”) for the vocal tasks were pre-recorded by one of the psychologists and played from a tablet computer ensuring standardized instruction over both experiments. The vocal tasks were recorded with the tablet computer’s internal microphone. Administration and recording were controlled by the application and facilitated the assessment procedure. To increase comparability, all recordings were sampled at 22.050 kHz and encoded with 16 Bit in the *wav* format.

### *Features*

Audio features were extracted directly and automatically from the audio signal. This form of speech analysis does not consider the semantic content of what a participant said, thus increasing the applicability of results in a clinical scenario, as no prior processing, such as transcription of what has been said, is required.

For each speech task (positive and negative story), features were extracted separately. The selection of vocal markers included standard acoustic measures and were guided by previous research on depression [29]. Overall, features were extracted from four different main areas: *prosodic*, relating to long-time variations in perceived stress and rhythm in speech. Prosodic features also measure alterations in per-

Table 1

Feature definition of acoustic markers. Name, definition, and intuition of features sorted by category is presented

| Category | Feature                         | Definition  | Intuition   |
|----------|---------------------------------|---|---|
| Prosodic | $F_0$                           | Mean, Max, Min, Range, Variance and Standard deviation of $F_0$   | Statistics over the perceived auditory pitch (speech melody)      |
|          | <i>Periodicity</i>              | Mean, Max and Min cross-correlation of speech signal  | Measure of the regularity of the speech signal                    |
| Formant  | $F_1$ – $F_3$                   | Mean and Variance of the first three formant frequencies  | Indicative of the class of speech sound                           |
|          | <i>Jitter</i>                   | Average absolute difference between consecutive signal periods, divided by the average period length        | Indicative for a lack of control for vibration of the vocal cords |
| Source   | <i>Shimmer</i>                  | Average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude | Indicative for reduction of glottal resistance                    |
|          | <i>Harmonics-to-Noise (HNR)</i> | Ratio between periodic components and aperiodic components comprising voiced speech                         | Measure of voice quality  |
|          | <i>Sounding segments</i>        | Mean, Max and Standard Deviation of sounding segment lengths determined based on intensity                  | Statistics over length of connected speech segments               |
|          | <i>Pause segments</i>           | Mean, Max and Standard deviation of silence segment lengths determined based on intensity                   | Statistics over length of continuous pause segments               |
| Temporal | <i>Duration</i>                 | Total length of recording   | Total length of recording   |
|          | <i>Speech duration</i>          | Total length of all sounding segments   | Amount of speech  |
|          | <i>Pause duration</i>           | Total length of all silence segments  | Amount of pause   |
|          | <i>Speech proportion</i>        | Ratio of <i>Speech duration</i> and <i>Duration</i>   | Proportion of recording participant is talking                    |
|          | <i>Speech rate</i>              | Ratio of number of syllables, detected using [38] and <i>Duration</i>                                       | Measure of information density                                    |
|          | <i>Articulation rate</i>        | Ratio of number of syllables, detected using [38] and <i>Speech duration</i>                                | Measure of speech tempo   |

sonal speaking style (e.g., perceived pitch, intonation of speech); *formant* features represent the dominant components of the speech spectrum and carry information about the acoustic resonance of the vocal tract and its use. These markers are often indicative of problems with articulatory coordination in speech motor control disorders (ref Sapir); *source* features relate to the source of voice production, the airflow through the glottal speech production system. These features operationalize irregularities in vocal fold movement (e.g., measures of voice quality); and *temporal* features include measures of speech proportion (e.g., length of pauses, length of speaking segments), the connectivity of speech segments and general speaking rate.

Table 1 gives a detailed overview, definition, and explanation of all extracted acoustic features. All features from the *temporal* category as well as  $F_0$  features were extracted using the Praat software (<http://www.fon.hum.uva.nl/praat/>). Jitter, Shimmer were determined using openSmile (<https://www.audeering.com/opensmile/>), a software for the extraction of vocal features. A Matlab script was used to extract HNR and statistics over the first three formants.

### Statistical analysis

All statistical analyses were run using R software version 3.4.0 (<https://www.r-project.org>). Because of the small sample size non-parametric tests were chosen. This study computed the Wilcoxon signed-rank and ranked-sum tests for dependent and independent sample testing respectively and Spearman's  $\rho$  for correlations. For the characterization of apathy, differences in acoustic measures are examined between the apathy and non-apathy group inside a gender. The goal being to find correlations between acoustic markers and the AI apathy sub-scales, as well as between acoustic markers, ultimately deriving properties of apathetic speech.

### Classification

Machine learning experiments were carried out to validate the diagnostic power of extracted markers. For this, classifiers were always trained within a gender (one classifier for males, one for females), to differentiate people with and without apathy.

As classifiers, simple Logistic Regression (LR) models implemented in the scikit-learn

Table 2  
Demographic data for population by gender and apathy

|                | Male         |                | Female       |                |
|----------------|--------------|----------------|--------------|----------------|
|                | N            | A              | N            | A              |
| N              | 12           | 12             | 18           | 18             |
| Age            | 78.25 (4.33) | 79.58 (5.45)   | 77.83 (6.12) | 79.50 (5.86)   |
| MMSE           | 22.66 (3.11) | 19.42 (4.17)   | 22.33 (4.02) | 19.56 (5.52)   |
| AI total       | 1.7 (1.23)   | 6.0*** (1.60)  | 0.56 (0.99)  | 5.33*** (1.97) |
| AI-Intr        | 0.75 (0.87)  | 2.42*** (0.90) | 0.17 (0.38)  | 2.33*** (0.91) |
| AI-Init        | 0.83 (0.94)  | 2.67*** (0.89) | 0.39 (0.69)  | 2.33*** (1.19) |
| AI-Affect      | 0.08 (0.29)  | 0.92** (0.90)  | 0.00 (0.00)  | 0.67** (1.08)  |
| NPI-Apathy     | 1.67 (2.01)  | 6.50*** (3.73) | 0.44 (0.70)  | 5.44*** (3.01) |
| NPI-Depression | 0.50 (0.90)  | 1.50 (2.68)    | 0.16 (0.51)  | 1.50 (2.91)    |
| NPI-Anxiety    | 1.50 (2.06)  | 2.75 (3.33)    | 0.94 (1.16)  | 3.11 (3.61)    |

Mean (standard deviation); Significant difference from the control population in a Wilcoxon-Mann-Whitney test are marked with \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . N, No Apathy; A, Apathy; MMSE, Mini-Mental State Examination; AI, Apathy Inventory; AI-Intr, AI domain Interest; AI-Init, AI domain Initiative; AI-Affect, AI domain affective; NPI, Neuropsychiatric Inventory; NPI-Apathy, NPI domain apathy; NPI-Depression, NPI domain depression; NPI-Anxiety, NPI domain anxiety.

(<https://scikit-learn.org/stable/>) framework were used. Linear models assign directly interpretable weights to each feature. Models using the L1 penalty (also referred to as lasso) are capable of performing implicit feature selection by reducing weights of unimportant features to zero. This was especially useful, since the number of used features is larger than the number of samples (see [47, p. 145] for more detail).

Because of the small data set, models were trained and evaluated in a leave-one-out cross-validation (LOO-CV) scenario. Here, all but one sample was used in training of the classifier and its performance is evaluated on the held out sample. This was repeated for all samples and performances were averaged.

Features were normalized using z-standardization based on the training set in each fold, excluding the held out sample. As a performance metric we report Area under the Curve (AUC) to be able to reason about possible specificity and sensitivity trade-offs.

## RESULTS

### Demographics

Demographic data is provided in (Table 2). After matching for MMSE and age, 24 male subjects and 36 female subjects were included in the final analysis and divided into equal groups of apathy and non-apathy subjects. No significant differences were present between the groups except for the results on the apathy scales.

### Correlation

Figure 1 presents Spearman correlation coefficients between extracted features and the AI sub-scales (i.e., *affective*, *interest*, *initiative*), split by gender. Only significant correlations are presented. The male population shows overall comparable correlations between the positive and the negative story. Generally, more significant correlations are observed for temporal features. In the positive story, correlations between these markers and all AI subdomains are present. Only a small negative correlation between  $F_0$ .

Range and the *affective* domain is observable ( $\rho = -0.47$ ). For the negative story, temporal features again dominate, while only showing correlations with the *interest* and *initiative* subdomains. Correlations with the *affective* domain are observed for both  $F_0$  Max ( $\rho = -0.61$ ) and  $F_0$  Range ( $\rho = -0.69$ ).

The female population shows more correlations in the positive story. Strong correlations are observed between all three subdomains and features relating to pause lengths. Features relating to sound length and speech tempo correlate significantly with the *interest* and *initiative* domain. In the negative story, nearly no correlations between temporal variables and any subdomain are present. Weak correlations are present between variables relating to mean Jitter (*affective*:  $\rho = 0.28$ ; *interest*:  $\rho = 0.29$ ), which is consistent with correlations in the positive story; minimum Shimmer (interest:  $\rho = -0.40$ ; initiative:  $\rho = -0.31$ ); and minimum (interest:  $\rho = 0.46$ ; initiative:  $\rho = 0.50$ ) and maximum Periodicity (interest:  $\rho = -0.47$ ; initiative:  $\rho = -0.41$ ).

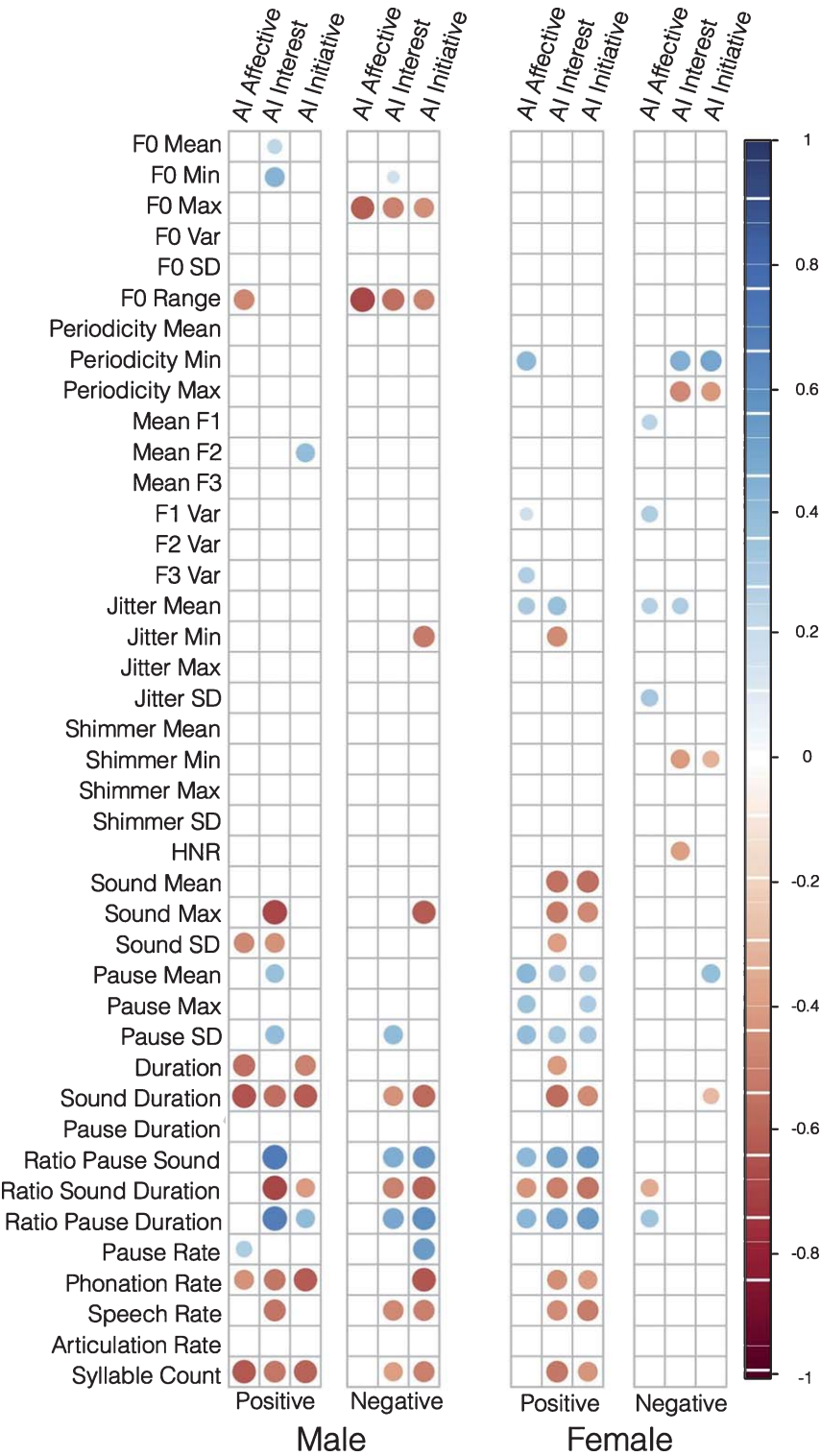


Fig. 1. Spearman correlation coefficient between features extracted from vocal tasks and AI subdomains. One correlation matrix is presented per speech task and gender. Only significant correlations ( $p < 0.01$ ) are displayed.

Table 3

Statistical group comparisons between non-apathetic and apathetic group using Kruskal-Wallis tests.

Features with  $p < 0.05$  are reported. Vocal task of origin,  $p$ -value, test statistic ( $\chi^2$ ), effect size ( $\rho$ ) and direction of effect in the apathetic group in comparison to the non-apathetic group are reported.

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

| A. Comparison for male population   |                      |              |                    |                    |           |
|-------------------------------------|----------------------|--------------|--------------------|--------------------|-----------|
| Origin                              | Feature              | Significance | Statistic $\chi^2$ | Effect size $\rho$ | Direction |
| Positive                            | Duration             | *            | 5.60               | 0.39               | ↓         |
|                                     | Ratio Pause Duration | **           | 6.75               | 0.43               | ↑         |
|                                     | Ratio Sound Duration | **           | 6.75               | 0.43               | ↓         |
|                                     | Ratio Pause Sound    | *            | 5.60               | 0.39               | ↑         |
|                                     | Sound Max            | *            | 6.45               | 0.42               | ↓         |
|                                     | Sound Mean           | *            | 5.33               | 0.38               | ↓         |
|                                     | Sound Duration       | ***          | 13.23              | 0.61               | ↓         |
|                                     | Pause Mean           | *            | 4.56               | 0.36               | ↑         |
|                                     | Syllable Count       | ***          | 11.81              | 0.57               | ↓         |
|                                     | Speech Rate          | **           | 9.36               | 0.51               | ↓         |
|                                     | Ratio Pause Duration | *            | 6.16               | 0.41               | ↑         |
|                                     | Ratio Sound Duration | *            | 6.16               | 0.41               | ↓         |
|                                     |                      |              |                    |                    |           |
| Negative                            | Ratio Pause Sound    | *            | 5.60               | 0.39               | ↑         |
|                                     | Sound Duration       | *            | 6.45               | 0.42               | ↓         |
|                                     | Sound Max            | *            | 4.08               | 0.34               | ↓         |
|                                     | Pause SD             | *            | 6.45               | 0.42               | ↑         |
|                                     | Pause Mean           | *            | 4.32               | 0.35               | ↑         |
|                                     | Pause Max            | *            | 4.08               | 0.37               | ↑         |
|                                     | Syllable Count       | *            | 3.85               | 0.33               | ↓         |
|                                     | Speech Rate          | *            | 5.88               | 0.40               | ↓         |
|                                     | $F_0$ Range          | **           | 9.72               | 0.52               | ↓         |
|                                     | $F_0$ Max            | **           | 9.36               | 0.51               | ↓         |
|                                     |                      |              |                    |                    |           |
|                                     |                      |              |                    |                    |           |
|                                     |                      |              |                    |                    |           |
| B. Comparison for female population |                      |              |                    |                    |           |
| Origin                              | Feature              | Significance | Statistic $\chi^2$ | Effect size $\rho$ | Direction |
| Positive                            | Ratio Pause Duration | **           | 10.62              | 0.54               | ↑         |
|                                     | Ratio Sound Duration | **           | 10.62              | 0.54               | ↓         |
|                                     | Ratio Pause Sound    | **           | 9.61               | 0.52               | ↑         |
|                                     | Sound Max            | **           | 6.73               | 0.43               | ↓         |
|                                     | Sound Mean           | **           | 8.29               | 0.48               | ↓         |
|                                     | Sound Duration       | **           | 8.66               | 0.49               | ↓         |
|                                     | Positive Pause Mean  | **           | 6.73               | 0.43               | ↑         |
|                                     | Pause Max            | *            | 5.48               | 0.39               | ↑         |
|                                     | Pause SD             | **           | 7.06               | 0.44               | ↑         |
|                                     | Syllable Count       | **           | 7.23               | 0.45               | ↓         |
|                                     | Speech Rate          | **           | 8.11               | 0.47               | ↓         |
|                                     | Jitter Mean          | *            | 5.33               | 0.38               | ↑         |
|                                     | HNR                  | **           | 6.73               | 0.43               | ↓         |
|                                     |                      |              |                    |                    |           |
| Negative                            | Periodicity Min      | **           | 8.11               | 0.48               | ↑         |
|                                     | Periodicity Max      | **           | 7.93               | 0.47               | ↓         |
|                                     | Jitter Min           | *            | 5.41               | 0.39               | ↓         |
|                                     | Jitter Mean          | *            | 5.05               | 0.37               | ↑         |
|                                     | Jitter SD            | *            | 5.33               | 0.38               | ↑         |
|                                     | HNR                  | **           | 9.42               | 0.51               | ↓         |

### Group comparison

Statistical comparisons between the apathetic and non-apathetic groups are presented in (Table 3A) for the male population and in (Table 3B) for the female population. Only significant values are reported.

Overall, features relating to temporal aspects of speech dominate. Some features show significant

differences regardless of gender (i.e., Speech Rate, Ratio Pause Duration, Ratio Sound Duration, Ratio Pause Sound, Sound, Max, Sound Duration), but for the female population only in the positive story. Males show significant differences in  $F_0$  Range and  $F_0$  Maximum in the negative story. Females show significant differences in HNR across both tasks. Females show differences in the negative story only in voice quality

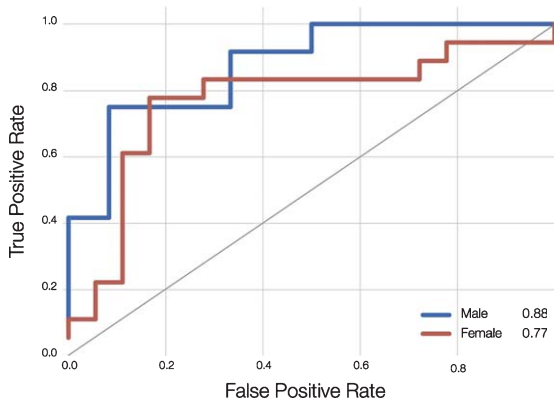


Fig. 2. Receiver Operator Curve (ROC) of classifiers trained to detect apathy from speech. The blue and red lines represent classifiers trained and evaluated on the male and female populations respectively. Area under the curve (AUC) is reported in the legend.

markers (Periodicity, Jitter, and HNR). For the male population, the largest effect in the positive story is the Sound Duration ( $\rho=0.61$ ) and for the F0 Range in the negative story ( $\rho=0.52$ ). For the females, the largest effects are in the Ratio Sound Duration for the positive story ( $\rho=0.54$ ) and the HNR for the negative story ( $\rho=0.51$ ). A table of the feature weights from the L1 regularized Logistic Regression models as well as spectrograms of non-apathy and apathy subjects during the positive and negative story telling task can be found in the Supplementary Material.

### Classification

Classification results are reported in (Fig. 2). AUC are far over the chance baseline of 0.5 for both male and female populations. Both AUC results are significantly better than the random chance baseline of 0.5.

The classifier trained on the male population achieves an AUC of 0.88 and the one trained on the female population an AUC of 0.77. The ROC visualizes a trade-off between sensitivity and (1 - specificity). For the male population, the classifier could be configured to achieve a good sensitivity of 0.91 and a reasonable specificity of 0.68. For the female population, a sensitivity of 0.85 and specificity of 0.72 can be configured.

## DISCUSSION

Early detection of apathy in older adults has reached high clinical relevance because of an increased risk of incidence of dementia and the dan-

ger to be easily overlooked by clinicians, which could lead to premature withdrawal from care [39]. The current study is the first one of its kind demonstrating clearly that certain paralinguistic features correlate significantly with levels of apathy severity. Thus, automatic speech analysis could be a promising new tool for its assessment.

Overall, the strongest correlations were found between the subdomains *interest* and *initiative* of the AI and temporal speech features. The *affective* subdomain, which represents the emotional blunting in apathy, was found to be more associated with prosodic speech features which is in line with previous findings on depressed speech with mainly prosodic speech abnormalities such as reduced pitch resulting often in a dull and ‘lifeless’ tone [29]. Similar observations were made in patients of this study with presence of emotional blunting. Thus, it seems that through speech features, distinct profiles can be characterized confirming what previous neuroimaging analyses revealed, namely that apathy is multidimensional and different subdomains are associated with different brain regions and circuits; the affective one with the ventral prefrontal cortex; the behavioral one with the basal ganglia; and the cognitive with the dorsomedial prefrontal cortex [40].

Overall, both males and females showed reduced reaction to the stimuli. Answers to the posed questions can be generally characterized by drastically shorter (lower sound duration) and slower (lower Speech Rate) speech. For the female population, a difference in voice quality (lower HNR) was obvious in both questions. Males suffering from apathy react less emotionally to the negative question as indicated by a lower variance of prosody (lower  $F_0$  Range). Interestingly, male and female subjects with apathy show different patterns in their speech features according to the type of free speech task. For males, significant differences between apathy and non-apathy subjects can be seen in temporal features for both the negative and positive story. Females show similar patterns in the positive story, but not in the negative one. Until today, no work on gender dependent symptoms of apathy has been found that could explain this pattern. Parts of this effect could be caused by the fact that men from this generation are in general less likely to talk enthusiastically about a positive event and show greater responses to threatening cues [41]. Gender differences in emotional processing and expressivity [42] as well as in emotional memory retrieval [43] could be another reason and should be further investigated, since current literature mostly focuses



on exploring age as a variable. Gender differences have been observed in brain activity during emotional tasks with primarily females recalling more autobiographical memories when it is of emotional content and cues are given verbally. It is possible that females in this study were more likely to be triggered to an emotional reaction when asked about a positive event and vice versa for males. Apathy might have an effect on this biased emotional memory retrieval. Hence, it can be assumed that the type of affective stimulus with which speech is being provoked might play a major role and might have to be adapted depending on a patient's gender.

Generally, when classifying between apathy and non-apathy subjects, features related to sound and pause segments seem to dominate with higher AUC results obtained for the male group. These features might have been particularly affected by the cognitive and behavioral aspects of apathy, which seem to be reflected in the general amount of speech produced. Recent findings suggesting that apathetic patients have decreased visual attentional bias for social stimuli compared with non-apathetic patients [20] might apply as well for speech production since it implies engagement in social interaction. Several reasons could explain these findings drawn from related studies on depression and negative symptoms in schizophrenia. This may be attributed to reduced muscle tension as well as impaired neuromuscular motor or articulatory coordination [24] caused potentially by alterations in the neurotransmitter system namely low norepinephrine and/or dopamine levels [43]. Changes in affective states can impact the phonation and articulation muscular systems via the somatic and autonomic nervous systems [44]. Commonly observed psychomotor retardation in apathy can lead to small disturbances in muscle tension which in turn can affect the speaker's speech pattern and, for instance, reduce pitch variability [45].

This study has some limitations. Instead of a standardized speech task (e.g., reading a text), this study relied on emotional questions for patients to elicit free speech. Although limiting the generalization to universal speech production, these tasks have proven effective in provoking speech that includes discriminative markers and are directly comparable to research about speech in depression [37]. Since patient data is always hard to acquire, the here presented sample is relatively small and future studies should strive to draw more conclusive evidence from larger datasets. Furthermore, this study considered three different statistical viewpoints of a single

dataset (i.e., group comparisons, correlations, and machine learning). Although these uni- and multivariate analysis are not independent, we ensured that results from one experiment did not directly influence another one (e.g., using the results of group comparisons in the classification experiment). Finally, due to the high number of correlated features and low sample size, no correction was applied to most significance tests, including the correlations with diagnostic scores.

Further work should examine what features in particular are predictive for apathy, how they relate to depression and how the two could be better discriminated. One potential solution could be to perform a semantic analysis of the content of speech to better differentiate apathy from depression and anxiety. Adding other additional measurements, for instance, of facial, head, or body movement dynamics, by the means of video might further improve accuracy. In the field of depression, research has demonstrated more powerful detection when applying a multi-modal audio-visual data fusion approach [46].

Nevertheless, it can be concluded that automatic speech analysis could become a promising new screening and assessment tool for follow-up measurements ('digital endpoints') in clinical trials of pharmacological and other interventions that aim to monitor apathy in patients.

## ACKNOWLEDGMENTS

The authors would like to thank all participants of this study. This research is part of the MNC3 program of the University Cote d'Azur IDEX Jedi. It was partially funded by the EIT Digital Well-being Activity 17074, ELEMENT, the University Cote d'Azur, and by the IA association and supported by the Edmond & Lily Safra Foundation and Institute Claude Pompidou.

Authors' disclosures available online (<https://www.j-alz.com/manuscript-disclosures/18-1033r2>).

## SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: <http://dx.doi.org/10.3233/JAD-181033>.

## REFERENCES

- [1] Mulin E, Leone E, Dujardin K, Delliaux M, Leentjens A, Nobili F, Dessi B, Tible O, Agüera-Ortiz L, Osorio RS,

- Yessavage J, Dachevsky D, Verhey F, Jentoft AJC, Blanc O, Llorca P, Robert PH (2011) Diagnostic criteria for apathy in clinical practice. *Int J Geriatr Psychiatry* **26**, 158–165.
- [2] Barch DM, Pagliaccio D, Luking KR (2016) Mechanisms underlying motivational deficits in psychopathology: Similarities and differences in depression and schizophrenia. *Curr Topics Behav Neurosci* **27**, 411–449.
- [3] Worthington A, Wood RL (2018) Apathy following traumatic brain injury: A review. *Neuropsychologia* **118**, 40–47.
- [4] Yuen, GS, Gunning-Dixon FM, Hoptman MJ, AbdelMalak B, McGovern AR, Seirup JK, Alexopoulos GS (2014) The salience network in the apathy of late-life depression. *Int J Geriatr Psychiatry* **29**, 1116–1124.
- [5] Yazbek H, Raffard S, Del-Monte J, Pupier F, Larue A, Boulenger JP, Gély-Nargeot MC, Capdevielle D (2014) L'apathie dans la schizophrénie: une revue clinique et critique de la question. *L'Encéphale* **40**, 231–239.
- [6] Aalten P, Verhey FR, Boziki M, Bullock R, Byrne EJ, Camus V, Caputo M, Collins D, De Deyn PP, Elina K, Frisoni G, Girtler N, Holmes C, Hurt C, Marriott A, Mecocci P, Nobili F, Ousset PJ, Reynish E, Salmon E, Tsolaki M, Vellas B, Robert PH (2007) Neuropsychiatric syndromes in dementia. Results from the European Alzheimer Disease Consortium: part I. *Dement Geriatr Cogn Disord* **24**, 457–463.
- [7] Pagonabarraga J, Kulisevsky J, Strafella A, Krack P (2015) Apathy in Parkinson's disease: clinical features, neural substrates, diagnosis, and treatment. *Lancet Neurol* **14**, 518–531.
- [8] Cathomas F, Hartmann MN, Seifritz E, Pryce CR and Kaiser S (2015) The translational study of apathy—an ecological approach. *Front Behav Neurosci* **9**, 241.
- [9] Merrilees J, Dowling G, Hubbard E, Mastick J, Ketelle R, Miller B (2013) Characterization of apathy in persons with frontotemporal dementia and the impact on family caregivers. *Alzheimer Dis Assoc Disord* **27**, 62–67.
- [10] Starkstein SE, Jorge R, Mizrahi R, Robinson RG (2006) A prospective longitudinal study of apathy in Alzheimer's disease. *J Neurol Neurosurg Psychiatry* **77**, 8–11.
- [11] Robert PH, Claret S, Benoit M, Koutaich J, Bertogliati C, Tible O, Caci H, Borg M, Brocker P, Bedoucha P (2002) The apathy inventory: assessment of apathy and awareness in Alzheimer's disease, Parkinson's disease and mild cognitive impairment. *Int J Geriatr Psychiatry* **17**, 1099–1105.
- [12] Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J (1994) The neuropsychiatric inventory comprehensive assessment of psychopathology in dementia. *Neurology* **44**, 2308–2308.
- [13] Clarke DE, Reekum RV, Simard M, Streiner DL, Freedman M, Conn D (2007) Apathy in dementia: An examination of the psychometric properties of the apathy evaluation scale. *J Neuropsychiatry Clin Neurosci* **19**, 57–64.
- [14] Robert P, Lanctôt KL, Agüera-Ortiz L, Aalten P, Bremond F, Defrancesco M, Hanon C, David R, Dubois B, Dujardin K, Husain M, König A, Levy R, Mantua V, Meulien D, Miller D, Moebius HJ, Rasmussen J, Robert G, Ruthirakuhan M, Stella F, Yessavage J, Zeghari R, Manera V (2018) Is it time to revise the diagnostic criteria for apathy in brain disorders? The 2018 international consensus group. *Eur Psychiatry* **54**, 71–76.
- [15] König A, Aalten P, Verhey F, Bensadoun G, Petit PD, Robert P, David R (2014) A review of current information and communication technologies: can they be used to assess apathy? *Int J Geriatr Psychiatry* **29**, 345–358.
- [16] Fitts W, Massimo L, Lim N, Grossman M, Dahodwala N (2016) Computerized assessment of goal-directed behavior in Parkinson's disease. *J Clin Exp Neuropsychol* **38**, 1015–1025.
- [17] Pessiglione M, Schmidt L, Draganski B, Kalisch R, Lau H, Dolan RJ, Frith CD (2007) How the brain translates money into force: a neuroimaging study of subliminal motivation. *Science* **316**, 904–906.
- [18] Kluge A, Kirschner M, Hager OM, Bischof M, Habermeyer B, Seifritz E, Walther S, Kaiser S (2018) Combining actigraphy, ecological momentary assessment and neuroimaging to study apathy in patients with schizophrenia. *Schizophr Res* **195**, 176–182.
- [19] David R, Mulin E, Friedman L, Duff FL, Cygankiewicz E, Deschaux O, Garcia R, Yessavage JA, Robert PH, Zeitzer JM (2012) Decreased daytime motor activity associated with apathy in Alzheimer disease: an actigraphic study. *Am J Geriatr Psychiatry* **20**, 806–814.
- [20] Chau SA, Chung J, Herrmann N, Eizenman M, Lanctôt KL (2016) Apathy and attentional biases in Alzheimer's disease. *J Alzheimers Dis* **51**, 837–846.
- [21] Faurholt-Jepsen M, Vinberg M, Frost M, Debel S, Margrethe Christensen E, Bardram JE, Kessing LV (2016) Behavioral activities collected through smartphones and the association with illness activity in bipolar disorder. *Int J Methods Psychiatr Res* **25**, 309–323.
- [22] König A, Linz N, Tröger J, Wolters M, Alexandersson J, Robert P (2018) Fully automatic analysis of semantic verbal fluency performance for the assessment of cognitive decline. *Dement Geriatr Cogn Disord* **45**, 198–209.
- [23] Bernardini F, Lunden A, Covington M, Broussard B, Halpern B, Alolayan Y, Crisafio A, Pauselli L, Balducci PM, Capulong L, Attademo L, Lucarini E, Salierno G, Natalicchi L, Quartesan R, Compton MT (2016) Associations of acoustically measured tongue/jaw movements and portion of time speaking with negative symptom severity in patients with schizophrenia in Italy and the United States. *Psychiatr Res* **239**, 253–258.
- [24] Covington MA, Lunden SA, Cristofaro SL, Wan CR, Bailey CT, Broussard B, Fogarty R, Johnson S, Zhang S, Compton MT (2012) Phonetic measures of reduced tongue movement correlate with negative symptom severity in hospitalized patients with first-episode schizophrenia-spectrum disorders. *Schizophr Res* **142**, 93–95.
- [25] Bortolon C, Macgregor A, Capdevielle D, Raffard S (2018) Apathy in schizophrenia: A review of neuropsychological and neuroanatomical studies. *Neuropsychologia* **118**, 22–33.
- [26] Blanchard JJ, Cohen AS (2006) The structure of negative symptoms within schizophrenia: implications for assessment. *Schizophr Bull* **32**, 238–245.
- [27] Strauss GP, Horan WP, Kirkpatrick B, Fischer BA, Keller WR, Miski P, Carpenter WT, Jr (2013) Deconstructing negative symptoms of schizophrenia: avolition-apathy and diminished expression clusters predict clinical presentation and functional outcome. *J Psychiatr Res* **47**, 783–790.
- [28] Compton MT, Lunden A, Cleary SD, Pauselli L, Alolayan Y, Halpern B, Broussard B, Crisafio A, Capulong L, Balducci PM, Bernardini F, Covington MA (2018) The aprosody of schizophrenia: Computationally derived acoustic phonetic underpinnings of monotone speech. *Schizophr Res* **197**, 392–399.
- [29] Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF (2015) A review of depression and suicide risk assessment using speech analysis. *Speech Commun* **71**, 10–49.

- [30] American Psychiatric Association (2013) *Diagnostic and statistical manual of mental disorders*, 5th ed. Washington, DC.
- [31] Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189–198.
- [32] Childers DG, Wu K (1991) Gender recognition from speech. Part II: Fine analysis. *J Acoust Soc Am* **90**, 1841–1856.
- [33] Heffernan K (2004) Evidence from HNR that/s/is a social marker of gender. *Toronto Working Papers in Linguistics*, **23**.
- [34] Wu K, Childers DG (1991) Gender recognition from speech. Part I: Coarse analysis. *J Acoust Soc Am* **90**, 1828–1840.
- [35] Linz N, Klinge X, Tröger J, Alexandersson J, Zeghari R, Robert P, König A (2018) Automatic detection of apathy using acoustic markers extracted from free emotional speech. In *Proceedings of the 2nd Workshop on AI for Ageing, Rehabilitation and Independent Assisted Living (ARIAL)* (Stockholm, Sweden, 2018), pp. 17–21.
- [36] Low, LSA, Maddage NC, Lech M, Sheeber LB, Allen NB (2011) Detection of clinical depression in adolescents' speech during family interactions. *IEEE Trans Biomed Eng* **58**, 574–586.
- [37] Cummins N, Sethu V, Epps J, Schnieder S, Krajewski J (2015) Analysis of acoustic space variability in speech affected by depression. *Speech Commun* **75**, 27–49.
- [38] De Jong, NH, Wempe T (2009) Praat script to detect syllable nuclei and measure speech rate automatically. *Behav Res Methods* **41**, 385–390.
- [39] van Dalen JW, van Wanrooij LL, Moll van Charante EP, Brayne C, van Gool WA, Richard ER (2018) Association of apathy with risk of incident dementia: A systematic review and meta-analysis. *JAMA Psychiatry* **75**, 1012–1021.
- [40] Kumfor F, Zhen A, Hodges JR, Piguet O, Irish M (2018) Apathy in Alzheimer's disease and frontotemporal dementia: Distinct clinical profiles and neural correlates. *Cortex* **103**, 350–359.
- [41] Kret ME, De Gelder B (2012) A review on sex differences in processing emotional signals. *Neuropsychologia* **50**, 1211–1221.
- [42] Deng Y, Chang L, Yang M, Huo M, Zhou R (2016) Gender differences in emotional response: Inconsistency between experience and expressivity. *PLoS One* **11**, 1–12.
- [43] Piefke M, Fink GR (2005) Recollections of one's own past: the effects of aging and gender on the neural mechanisms of episodic autobiographical memory. *Anat Embryol* **210**, 497–512.
- [44] Mitchell RA, Herrmann N, Lanctôt KL (2010) The role of dopamine in symptoms and treatment of apathy in Alzheimer's disease. *CNS Neurosci Ther* **17**, 411–427.
- [45] Scherer KR (1986) Vocal affect expression: A review and a model for future research. *Psychol Bull* **99**, 143–165.
- [46] Horwitz R, Quatieri TF, Helfer BS, Yu B, Williamson, JR, Mundt J (2013) On the relative importance of vocal source, system, and prosody in human depression. In *2013 IEEE International Conference on Body Sensor Networks*, pp. 1–6.
- [47] Dibeklioglu H, Hammal Z, Yang Y, Cohn JF (2015) Multimodal detection of depression in clinical interviews. In *Proceedings of the ACM on International Conference on Multimodal Interaction* (New York, USA, 2015), ICMI '15, ACM, pp. 307–310.
- [48] Bishop C (2006) *Pattern Recognition and Machine Learning* (Information Science and Statistics). Springer-Verlag, Berlin, Heidelberg.